Uhe, P. F., Mitchell, D. M., Bates, P. D., Allen, M., Betts, R. A., Huntingford, C., King, A., Sanderson, B., & Shiogama, H. (2021). Method-uncertainty is essential for reliable confidence statements of precipitation projections. *Journal of Climate*, *34*(3), 1227–1240. https://doi.org/10.1175/JCLI-D-20-0289.1

Method-uncertainty is essential for reliable confidence statements of precipitation projections

Peter Uhe*, Dann Mitchell and Paul D. Bates

School of Geographical Sciences, University of Bristol, Bristol, UK

Myles R. Allen

Environmental Change Institute, University of Oxford, Oxford, Oxfordshire, UK

Richard A. Betts

Met Office Hadley Centre, Exeter, Devon, UK. Global Systems Institute, University of Exeter, MetOffice, Exeter, Devon, UK

Chris Huntingford

Centre for Ecology and Hydrology, Wallingford, Oxfordshire, UK

Andrew D. King

ARC Centre of Excellence for Climate Extremes, School of Earth Sciences, University of Melbourne, Melbourne, Victoria, Australia

Benjamin M. Sanderson

European Center for Research and Advanced Training in Scientific Computing, Toulouse, France

Hideo Shiogama

18 *National Institute for Environmental Studies, Tsukuba, Ibaraki 305-0053, Japan*

19 **Corresponding author*: Peter Uhe, Peter.Uhe@bristol.ac.uk

ABSTRACT

Precipitation events cause disruption around the world and will be altered by climate change. However, different climate modeling approaches can result in different future precipitation projections. The corresponding 'method-uncertainty' is rarely explicitly calculated in climate impact studies and major reports, but can substantially change estimated precipitation changes. A comparison across five commonly-used modeling activities shows that for changes in mean precipitation, less than half the regions analyzed had significant changes between the present climate and 1.5°C global warming for the majority of modeling activities. This increases to just over half the regions for changes between present climate and 2°C global warming. There is much higher confidence in changes in maximum 1-day precipitation than in mean precipitation, indicating the robust influence of thermodynamics in the climate change effect on extremes. We also find that none of the modeling activities capture the full range of estimates from the other methods in all regions. Our results serve as an uncertainty map to help interpret which regions require a multi-method approach. Our analysis highlights the risk of over-reliance on any single modeling activity and the need for confidence statements in major synthesis reports to reflect this 'method-uncertainty'. Considering multiple sources of climate projections should reduce the risks of policymakers being unprepared for impacts of warmer climates compared to using single-method projections to make decisions.

1. Introduction

Understanding future precipitation changes in a warming world is critical to empower communities to make informed decisions around adaptation or climate related policy. Precipitation provides drinking water, is relied on for agriculture and used in many sectors of industry, so changes in water availability need to be understood to make the most of this limited resource. Droughts cause severe strain on people and ecosystems. Storms and extreme rainfall events also cause flooding and destruction. Worldwide, flooding affects more people than any other natural disaster (Wallemacq and House 2018).

Unfortunately, given the importance of precipitation for daily life, future changes in precipitation are much less certain than temperature changes (Collins et al. 2013; Tebaldi et al. 2011). In this study we look at low levels of global warming, in particular 1.5°C and 2°C, which are relevant to the Paris Agreement and associated policy decisions. A challenge relating to these levels of warming is that the signal of precipitation changes can be difficult to distinguish from the noise as they are often small relative to internal variability (Hawkins and Sutton 2011), and require larger ensemble sizes to detect than temperature trends (Deser et al. 2012). There are non-linear effects in the climate system and differences between transient and equilibrium climate response, so changes based on higher levels of warming cannot simply be used to estimate impacts for 1.5°C and 2°C (Good et al. 2016; Mitchell et al. 2016). Furthermore, precipitation events are tightly connected to atmospheric and ocean dynamics and changes are seasonally dependent so interpreting changes in precipitation and their impacts requires careful analysis.

The most common approach when investigating future changes of precipitation is to use general circulation models (GCMs) which dynamically simulate the physics of the atmosphere and ocean. Different GCMs use varying representations of the physics, so model inter-comparison projects

(MIPs) are frequently used to provide a range of different possible futures. The MIPs used in this study (also referred to as modeling activities) are the Coupled Modeling Intercomparison Project, Phase 5 (CMIP5) (Taylor et al. 2012), the Coupled Modeling Intercomparison Project, Phase 6 (CMIP6) (Eyring et al. 2016; O'Neill et al. 2016), the Half a degree Additional warming, Prognosis and Projected Impacts project (HAPPI) (Mitchell et al. 2017), the 2018 UK Climate Projections (UKCP18) (Murphy et al. 2019) and the High-End cLimate Impacts and eXtremes project (HELIX) (Wyser et al. 2017). MIPs provide a common experimental protocol under which multiple modeling groups run simulations to produce multi-model ensembles of climate projections. The use of MIPs has been a successful approach, and Fig. 1 shows that around half of the impact studies in the Intergovernmental Panel on Climate Change (IPCC) Special report on Global Warming of 1.5°C (IPCC 2018) result directly from one of these MIPs.

In producing their assessment reports, the IPCC strives to compile information across all the available literature. However, it relies heavily on using the latest modeling inter-comparison project to determine the likelihood of changes in climate. For example in the IPCC Fifth Assessment Report (AR5), the CMIP5 results were compared with the previous activity (CMIP3) to see how they differ. However the keynote plots in the IPCC 'Atlas of Global and Regional Climate Projections' were solely from the CMIP5 ensemble. In the coming years, there will be a strong focus on analyzing the latest results from CMIP6 which will contribute to the IPCC Sixth Assessment Report (AR6). CMIP6 has a broad sample of current model diversity with a generally higher model complexity than CMIP5, so there are many benefits to using this new resource. However, single MIPs such as CMIP5 can under-estimate the possible range of future climate change (Deser et al. 2020). On the other hand, GCMs have a range of climate sensitivities to greenhouse-gas forcing (Sherwood et al. 2014) and CMIP6 is known to have a large proportion of high climate sensitivity models (Zelinka et al. 2020), which may overestimate the upper bound of warming (Tokarska et al. 2020). So

especially in regions with low confidence in precipitation change, it could be counter-productive to disregard the huge resource of previous climate model results and focus on CMIP6 alone.

Within each MIP, a common experimental design is used. However different experimental designs can lead to differing impacts of 1.5°C warming, related to factors such as the rate of global warming and the aerosol forcing relative to greenhouse gas forcing (Seneviratne et al. 2018; King et al. 2018). The large CMIP5 and CMIP6 activities use a number of different emissions scenarios, so do include a measure of scenario uncertainty. However there are other uncertainties relating to experimental design, such as the use of high-resolution cloud or convection resolving models compared to models which parameterize these processes, or the inclusion of carbon-cycle feedbacks compared to prescribed greenhouse-gas forcing. The differences in climate response between transient and equilibrium climate are also difficult to diagnose using traditional scenario-based MIPs, which produces another source of experimental design uncertainty that is relevant to policy decisions. Our study aims to take the comprehensive approach of analyzing results from MIPs which use different modeling approaches. Here we examine uncertainty, not just due to different emission pathways in a single MIP, but differing experimental setups in different MIPs.

There is a risk that relying on a single MIP may result in over-confidence in climate projections by missing some uncertainty due to experimental design. In addition, considering different emissions pathways at lower levels of warming can give different precipitation changes (Mitchell et al. 2016). On the other hand, comparisons between CMIP3 and CMIP5 high emissions pathways show consistent changes in seasonal precipitation (Knutti and Sedláček 2013), which increases the confidence in those results. Hence determining agreement in precipitation projections can enhance (where they agree) or reduce (where they disagree) our confidence in the individual projections.

In Fig. 1, only a very small proportion of studies considered a combination of approaches to obtain multiple lines of evidence regarding future changes. Combining large multi-model ensembles of

simulations with differing experimental design and skill at representing the current climate is not straightforward. We note that it is not always clear that improved model skill in the present day results in improved future projections (Knutti et al. 2010). However, there is ongoing work regarding weighting simulations depending on their representation of relevant climate phenomena or relation to other simulations (e.g., Sanderson et al. 2017a; Merrifield et al. 2020; Brunner et al. 2020). This has the potential to constrain the likely range of future projections for example by down-weighting high climate sensitivity models which give poor performance over the historical period.

This study focuses on the agreement across multiple modeling activities, of estimates of precipitation change at specific levels of global warming (e.g. 1.5°C and 2°C). We compare changes in yearly mean precipitation and the yearly maximum of daily precipitation ('extreme precipitation'). We use averages over land of updated reference regions created for the IPCC AR6 (Iturbide et al. 2020, see Fig. S1) to investigate different regional signals. Time-slices of transient simulations are used to examine specific levels of global warming. We consider each of the MIPs used in this study as providing plausible representations of future climate and do not weight any one higher than the others. This is reasonable given their individual use in different analyses of projected precipitation change.

We firstly show the agreement in sign of significant changes to 1.5°C and 2°C warming across the five climate modeling activities. This approach identifies regions where modeling activities agree in a significant change, and regions in which the change is more uncertain. The significance is determined from the 5-95% confidence intervals of the 'central estimates' calculated for each MIP. The central estimate is calculated by combining the model estimates within each MIP, taking into account the model spread and sampling uncertainty for each model. A combined central estimate for results across the MIPs is also calculated.

In addition to showing the combined changes and whether the changes are significant, we also consider uncertainty in each of the modeling activities' results and the combined central estimate. The magnitude of uncertainty bounds and the extent of overlap between uncertainty estimates is explored. Furthermore, to dig deeper into uncertainty due to experimental design, we undertake comparisons between changes calculated for different experimental designs or scenarios. This is done using two individual models that each have large ensembles of simulations, as well as by comparing different scenarios within the CMIP5 and CMIP6 activities.

This analysis illustrates the potential of combining the agreement across different modeling activities with a more detailed examination of experimental design using single model large ensembles. This approach provides a fuller picture of the 'method-uncertainty' in these climate modeling activities. This is something that is difficult to quantify but is essential to address, especially in regions where the changes are not as clear as a single modeling activity would indicate.

2. Materials and Methods

Methods for analyzing results from GCM simulations are presented below. Information about the specific climate model data-sets is given in the Appendix.

a. Climate indices and regions

For this analysis, we focus on two precipitation indices. We use the annual mean precipitation (referred to as 'mean precipitation') and the yearly maximum of daily precipitation (referred to as 'extreme precipitation'). The mean precipitation is used to indicate whether there is a change in the total amount of precipitation over a region. The extreme precipitation index is used to indicate whether there will be a change in the magnitude of precipitation in heavy rainfall events or storms.

When looking at impacts in specific sectors and local scales, indices that capture seasonality are also very useful, but we chose these two indices as they are widely applicable on a global scale.

When calculating the changes in precipitation between different specific warming levels, we focus on the percentage changes, to show the changes relative to the model climatology. This gives a normalized metric of changes to reflect that a mean change, for example 0.2 mm/day, in a low rainfall area is likely to make a larger impact than the same change in a very high rainfall region. The use of relative changes does mean that in the presence of model biases, the same absolute change in precipitation will appear as different percentage changes. In addition, in areas of very low precipitation, showing percentage changes of relative changes may over-emphasize small changes in precipitation. To support these analyses, we additionally show results of absolute changes (in mm/day) as supplementary material.

For analysis of changes, region definitions were used as per (Iturbide et al. 2020). These regions were developed as an update to regions used in the IPCC AR5 and the IPCC SREX report, using smaller regions in some parts of the world to achieve better climatic consistency within each region. A map of these regions is shown in Fig. S1, labeled with the acronyms used for each region. The precipitation indices were first averaged over these regions before calculation of changes. Note that in the regions analyzed here, averages were calculated over land points only.

b. Extracting 1.5°C and 2°C time-slices

Transient GCM experiments are designed around simulations of the historical period then continuing into the future using scenarios representing different emissions pathways. From these simulations, we can then determine the climate state when these scenarios reach different levels of global warming. In this study, we use a commonly used approach of selecting time-slices (King et al. 2017; James et al. 2017). This approach does have the limitation that climate from a transient

climate simulation can differ from simulations stabilized at the same level of warming due to effects that lag behind the warming of the atmosphere (e.g. ocean circulation and sea-level rise) (Manabe et al. 1991; Held et al. 2010). The alternative is to compute targeted simulations that stabilize at each specific level of warming, but this has only been done in a few cases (e.g. Sanderson et al. 2017b), so using time slices of transient simulations is still a widely used method.

Firstly, a baseline is chosen as the start of the historical period (e.g. 1861-1900), to calculate the pre-industrial reference temperature. Then 21-year time-slices are chosen for the first period that has the global mean temperature averaged over the time-slice reaching the specific warming levels of 1.5 and 2°C relative to the baseline. For current climate, time-slices for the warming level of 0.9°C are used to match observed warming to 2010. This is done, rather than taking a fixed time period, to keep the warming between the current and 1.5°C time-slice consistent, and thereby accounting for the variation in climate sensitivities between models. We note that this will inevitably result in there being different aerosol forcings between models in each of the current, 1.5°C and 2°C warmer worlds. As the historical simulations are not necessarily long enough to capture our current climate period (in CMIP5 they finish in 2005), they are extended by future scenario simulations where necessary. When more than one future scenario was available, the highest emission scenario was used to extend the historical simulation for the current climate period. This prevents low climate sensitivity models (which reach 0.9°C later) from having current climate time-slices as far into the future scenarios as would be the case using low emissions scenarios. Note, the current climate time-slices are referred to as 'Hist' in some figures.

For CMIP5 and CMIP6, simulations from all future scenarios available are included in the analysis to maximize the number of samples. The exception to this is the results for section 3c, where the changes calculated using low and high warming scenarios were compared.

In this study we aim to keep the methodology of extracting specific levels of warming as consistent as possible. However, different experimental designs do mean that the time-slices need to be calculated in different ways in some cases. These differences are described in the Appendix for each data-set where relevant.

c. Statistical Analysis

To estimate the change in a particular variable between two time-slices, all of the years in each time-slice for each model are pooled together. Then the ensemble mean response is determined based on all years of data for that particular model. The uncertainty range in the mean response is determined by randomly re-sampling each distribution with replacement 1000 times and calculating the mean response from each sample. The 5-95th percentile range of the samples then gives the sampling uncertainty in the mean change.

When determining the significance of multi-model changes, for example in the IPCC report, it is common practice to use significance tests to determine whether changes are distinguishable from natural variability alongside thresholds for the proportion of models agreeing on the sign of the change (e.g., Tebaldi et al. 2011). However, these type of approaches do not provide a confidence interval around the multi-model change, making it difficult to combine uncertainty estimates of different multi-model data-sets together.

Here, to combine each of the model estimates into to a multi model summary or 'central estimate', we use the random-effects meta-analysis method (Cochran 1937; DerSimonian and Laird 1986). This methodology is commonly used in clinical studies to combine central estimates and uncertainty ranges of different studies together and was applied to climate models in Uhe et al. (2019). Such a statistical approach takes into account both the sampling uncertainty from random re-sampling (s_i) and the model spread (σ) which is taken as the standard deviation of the central estimates. From

these quantities, a a combined central estimate of change and an estimate in the uncertainty in that value are derived.

The multi-model central estimate μ and its standard error δ are given by the following equations:

$$\mu = \sum_i (w_i.\mu_i) / \sum_i w_i \tag{1}$$

$$\delta = (\sum_i w_i)^{-1/2} \tag{2}$$

$$w_i = (s_i^2 + \sigma^2)^{-1} \tag{3}$$

where w_i are the weights given to each of the model estimates in the calculation of the central estimate. The 5-95% confidence interval is calculated as $\mu \pm 1.6\delta$, assuming normally distributed values.

This calculation of central estimates is applied to combine different model estimates for each of the MIPs, and also finally to combine the central estimates of each MIP into an overall 'Combined central estimate'. The changes are referred to as statistically significant if the 5-95% confidence interval does not include zero.

3. Results

a. Regional changes and agreement

To evaluate the confidence in large scale patterns of precipitation changes, we use agreement between climate modeling activities. Fig. 2 shows the agreement of changes between current climate and 1.5°C or 2°C, across our five MIPs, for mean and extreme precipitation. Agreement here is represented by the number of modeling activities which show a significant change, i.e. the 5–95% confidence interval not including zero.

In Fig. 2, regions are marked with hatching where there are conflicting but significant changes from two different MIPs. Encouragingly, this shows that there are only a few regions for mean precipitation (North Central-America, Sahara, South Eastern-Africa and southern South-America) where two different modeling activities have significant changes with opposite signs, between current climate and 1.5°C . For the changes to 2°C , this is reduced to just southern South-America.

CMIP6 is the latest MIP, using current state-of-the-art climate models, and will underpin most of the conclusions described in the IPCC AR6. For this reason, we highlight regions where CMIP6 does not agree in the significance of the changes with the majority of other modeling activities. In Fig. 2, bold outlines indicate where CMIP6 gives a different sign or significance in the changes to three of the other four modeling activities. This identifies vulnerable regions such as some parts of South America or Africa, where using information from CMIP6 alone may misrepresent our confidence in the precipitation changes to 1.5°C of global warming. We note that these are not indicating that the using CMIP6 results in a different sign to significant changes given by other MIPs, rather it may give a significant change where most other MIPs show only insignificant changes, or vice versa. However, this is still an important point as it is relevant to the confidence statements produced by the IPCC (or other major reports), which may be considered by decision makers regarding climate change planning.

Fig. 3 shows the percentage changes in mean and extreme precipitation, from the combined central estimate of the five modeling activities. To highlight the confident changes, regions where the combined central estimate gives a significant change are marked with a bold border in Fig. 3. We additionally include the same changes, but calculated in mm/day in Fig. S2. For breakdown by modeling activity, Figs S3 and S4 show the changes and the significance of the central estimates for each MIP, in percentage change and mm/day respectively.

From Figs 2 and 3, we see that the precipitation changes in North America and Eurasia show the strongest agreement, especially at the lower warming level of 1.5°C. For changes in the southern hemisphere and some equatorial regions, there is often less agreement. Hence, in these regions, the use of a single modeling activity (as most studies have done) risks creating false confidence in the changes.

Changes in extreme precipitation show a large amount of agreement. At 2°C warming, the majority of modeling activities show confident changes in almost all regions (except the Sahara and Caribbean regions). This higher confidence in extreme precipitation has been reported previously (Allen and Ingram 2002; Fischer et al. 2014; Pendergrass et al. 2015). This is due to thermodynamics dominating extreme precipitation changes, while mean precipitation will be more strongly influenced by dynamical i.e. atmospheric circulation changes, which have less certainty and more disagreement. We also note that there are increases in extreme precipitation in regions which show drying changes in the mean precipitation. This increase in extreme precipitation could be part of the source of uncertainty in mean precipitation drying, due to the extreme precipitation contributing different fractions of the total precipitation in different models.

The level of agreement between the different modeling activities in the precipitation changes is also strongly connected to the strength of the changes. Fig. S5 shows the signal to noise ratio for each of the modeling activities, where the noise represents the magnitude of the 5–95% confidence intervals. Here we see that the areas which have the highest agreement also have the strongest signal to noise ratio. A useful metric to measure of magnitude of changes is the internal variability of the system, and Fig. S6 shows normalized changes, representing the amount of the change relative to the variability simulated by each model. This highlights that at these small levels of global warming, many of the changes are smaller than the year-to-year variability, however can be detected confidently by using the large number of samples in these MIPs.

In addition to the agreement in the sign of the precipitation changes, it is relevant to understand whether the uncertainty range in changes using each modeling activity overlap. For this, we consider the changes in mean precipitation between current climate and 1.5°C warming. Figure 4 shows the amount of overlap between the confidence interval for each MIP and the confidence intervals calculated for the combined central estimate of the other MIPs. In nearly all regions there is some overlap between the modeling activities, so it is rare for the central estimates of each modeling activity to completely disagree. We note that in Fig. 4, a value of 100% does not necessarily indicate perfect agreement. Instead, it can reflect a larger uncertainty range in the changes for a given MIP, which encompasses the combined central estimate for the other MIPs. Part of this may be due to the nature of the combined central estimate, which can have a smaller uncertainty range if the models are in agreement, reflecting the greater number of samples included. Figures S7–9 show similar results for extreme precipitation and 2°C warming.

We highlight that the HAPPI activity shows more regions where the central estimate disagrees with the other modeling activities. This may be partly due to the large initial condition ensembles within HAPPI resulting in smaller uncertainty bounds, but at the same time not including uncertainty in the ocean and sea-ice responses, hence giving overconfident estimates. HAPPI and HELIX also exhibit a tendency to give different results in some northern regions, which may indicate an influence from the prescribed sea-ice that is used in their atmosphere-only simulations. Looking at Fig. 4 and Figs S7–9, there is no activity that agrees with the combined result from the other activities in all cases. This finding provides substantial support to the benefit of considering a range of modeling activities.

b. Partitioning of uncertainty

When considering the confidence of a particular model result, understanding the source of uncertainties can be highly illustrative. We consider three types of uncertainty: sampling uncertainty, inter-model uncertainty and experimental design uncertainty (the latter of which is considered in detail in section 3c).

We consider firstly sampling uncertainty within a single model projection, calculated as per section 2c. This uncertainty is related to the internal variability in the climate system and the number of years of simulation included in the sample. To reduce the uncertainty in a single model response, modeling centers generate ensembles of simulations, usually produced by initial condition or physics parameter perturbations. We also consider the uncertainty in the central estimates for each MIP. We note that the central estimate uncertainty is not an independent quantity, but is calculated based on the confidence intervals of each model, as well as the spread of model changes. We finally consider the combined central estimate uncertainty.

Figure 5 shows these different quantities of uncertainty in the combined projections to 1.5°C and 2°C. Four regions are shown as illustrative examples. With regards to sampling uncertainty (i.e. single model uncertainty), HAPPI, which uses large ensembles has a much smaller sampling uncertainty than CMIP5 and CMIP6, which mostly have fewer than three historical simulations per model (see Tables S1–3 for ensemble sizes). HAPPI simulations also use atmosphere-only models, forced by a single set of prescribed sea-surface temperatures, so HAPPI may represent a smaller range of possible futures compared to the full spread of coupled ocean-atmosphere models.

In Fig. 5, the combined central estimate uncertainty is at the lower end of the single MIP central estimate uncertainties. This finding is a result of the construction of the central estimate 'narrowing in' on the most plausible response as more samples are available. We note though, that this is a

purely statistical approach to determining the uncertainty range. In terms of ability to model the climate system, outlier models may be just as plausible, despite lying outside our central estimate uncertainty. Other things that could be considered are model inter-dependencies e.g. different models sharing code or components (Knutti et al. 2013).

We also note that in a commonly pictured view of model uncertainty (Hawkins and Sutton 2009), the model uncertainty in a given variable increases over simulated future times. This increasing spread is partly because different models have different climate sensitivities and therefore warm at different rates. However, as we are examining model projections at specific levels of warming, any first order differences due to climate sensitivity will not be included in our uncertainty estimates. Lehner et al. (2020) showed that model uncertainty for global mean precipitation also increases with global warming, with small differences between CMIP5 and CMIP6, but here we look at uncertainty for a few specific regions. In Fig. 5, we show that while the central estimate uncertainty does generally increase, there are cases where it stays constant or decreases as global warming increases, for example the HAPPI projection of mean precipitation over the Mediterranean or the UKCP18 projections of extreme precipitation over western Central Europe. Where the single model (sampling) uncertainty does not show substantial changes, we expect the changes in central estimate uncertainty to relate to model uncertainty. In other regions shown here, the uncertainty is similar or increases as warming rises from 1.5°C to 2°C, but this highlights that the use of specific levels of warming can constrain the uncertainty.

c. Differences in experimental design and scenarios

In addition to the uncertainty at the model or MIP level, there is uncertainty due to the experimental design of each modeling activity. The previous section considered the uncertainty across the MIPs, however this is not the same as the experimental design uncertainty. As each of the

MIPs use different models (and different generations of models), it is not possible to formally connect the multi-MIP spread directly to the experimental design. However, the experimental design uncertainty can be related to the choice of scenario and forcing data-sets used to run the future projections. The experimental design uncertainties may also involve more structural differences for example the use of atmosphere-only compared to coupled ocean-atmosphere models, or the choice of using a dynamic carbon cycle with emissions prescribed rather than GHG concentrations.

To isolate the influence of experimental design on the future projections, one approach is to use single model large ensembles. Where these large ensembles have produced simulations using multiple modeling protocols, we can compare their responses at specific levels of warming. For this analysis we have used the CanESM2 large ensemble (Kirchmeier-Young et al. 2017) using the RCP8.5 scenario from CMIP5, and compared it with the CanAM4 (the atmospheric component of the CanESM2 model) simulations produced using the HAPPI scenarios. Secondly, we have compared the CESM large ensemble (Kay et al. 2015) using the RCP8.5 scenario from CMIP5, with the CESM low warming simulations (LowWarm) using emissions pathways designed to stabilize at 1.5 or 2°C (Sanderson et al. 2017b).

Figure 6 shows the comparison between the experimental designs over different regions. Differences for CanESM2 are shown in the upper panel and differences for CESM are shown in the lower panel. The differences shown are for mean precipitation, comparing the percentage changes from current climate to 1.5°C between the two experimental designs. Regions that are outlined in bold are where the significance or sign of the change is different between experimental designs. For both models, there is a clear difference over the Americas where the stabilized scenario (HAPPI or LowWarm) becomes wetter relative to the transient RCP8.5 simulations. Similar differences are seen over Asia, although with less consistency. An opposite trend is seen over the Northern and Eastern African regions, and parts of Australia.

Two factors causing a difference between the stabilized and transient simulations, are the differences in non-greenhouse gases such as anthropogenic aerosols, and the differences in the in land-sea contrast driven by the land warming faster than the ocean. Anthropogenic aerosols are projected to be significantly reduced by the end of the 21st century which is reflected in the stabilized scenarios. The transient simulations, however, may pass the 1.5°C temperature threshold before the mid-21st century, and so will have significantly higher modeled aerosol loads. This may be reflected in the relatively strong differences in East Asia in Fig. 6, particularly for CESM. We note that models with different representations of aerosols will give differing changes, which may be a source of model uncertainty in the multi-model analysis, for areas of high aerosol forcing.

We investigate spatial patterns of changes over the oceans in Fig. S10, which is as per Fig. 6 but instead showing model grid-cells rather than regional averages. There are strong positive precipitation anomalies on the Pacific equator indicating differences in the Pacific Intertropical Convergence Zone between stabilized and transient simulations. This could be related to differences in the north-south warming contrast between the experiments. Also in Fig. S10, there is a pattern of wetting over the Atlantic Ocean and drying in the north of Africa in the stabilized experiments relative to RCP8.5. This may be due to the land-sea contrast from the Sahara region warming much faster than the Atlantic Ocean in the transient simulations. In the stabilized experiments, the Atlantic Ocean warming may catch up, causing this difference.

We additionally look into the differences between low and high warming scenarios for the CMIP5 and CMIP6 ensembles in Fig. 7. These do show some regions where the significance of the change is different between scenarios. Differences here are important when considering the implications of following a low emissions pathway, and in these bold region (covering large parts of America and Africa), careful evaluation of the different scenarios should be performed separately. The differences here are smaller than the single model differences in Fig. 6, probably due to differing

responses in models within CMIP5 and CMIP6. There are also only a few regions which show notable changes which are consistent between CMIP5 and CMIP6, e.g. parts of in central America, central Africa and New Zealand. Other regions have small differences or are not consistent between CMIP5 and CMIP6.

Finally, the smaller difference in these scenarios for CMIP5 and CMIP6 in many regions may be attributable to the low warming amount of 1.5°C. The CMIP models are not in equilibrium by the time they reach 1.5°C of global warming, even for the low warming scenarios, so the comparison in Fig. 7 does not clearly represent an equilibrium vs transient climate in the same way as in Fig. 6. In addition, the differing model responses and the small number of ensemble members makes it difficult to identify any signal due to scenario differences for this analysis. Again, this shows the value of the single model large ensembles used above.

4. Discussion and Conclusions

Uncertainty arising from differences between climate modeling activities is often ignored in climate change studies and reports. As these studies form the basis for climate change policy, 'method-uncertainty' is essential for reliable confidence statements of precipitation change.

This article presents a statistical method to combine projected estimates of change from multiple modeling inter-comparison projects. This involves producing a 5-95% confidence interval, which is used to determine a statistically significant change. This approach has the advantage that the uncertainty range is determined from the sampling uncertainty of each model and the spread across different model changes, and does not rely on arbitrary thresholds such as percentage of models that agree. We argue that using such a method and evaluating the agreement between modeling intercomparison projects and the combined central estimate from a range of different projects gives a quantification of the method-uncertainty.

This study shows the agreement in precipitation changes between five different modeling activities. For mean precipitation, just over half of the regions have a significant change in the majority of modeling activities for changes to 2°C. In contrast, for increases in extreme precipitation there are significant changes for the majority of the MIPs almost everywhere by 2°C warming. Regarding the magnitude of possible changes, we also show that there is no single modeling activity which captures the full range of changes estimated by the other MIPs in all cases.

We note that drying is less confidently predicted than the wetting. Drying in mean precipitation can occur while the extreme precipitation is increasing, which may obscure some of the signal. Another consideration, is that the region definitions themselves may not enable identification of drying on smaller spatial scales. The nature of precipitation as a positive quantity also sets an upper bound on the possible amount of drying, particularly in already dry regions, which may cause the wetting changes to overcome drying over larger regional averages. It is also possible that the location of the drying regions is slightly different between models, and calculating a multi-model mean results in a loss of signal (e.g., Knutti et al. 2010). Nonetheless, model spread and disagreement across modeling activities need to be taken into account when evaluating risks associated with these changes. More detailed seasonal level analysis of these regions also will supplement these findings.

Furthermore, it is necessary to understand the sources of uncertainty in each of the modeling activities, and the method they use to determine future changes in climate. The CMIP5 and CMIP6 projects provide a large structural sample by including many coupled ocean-atmosphere models, but have limited numbers of simulations per model. The HAPPI project contains a range of models and has large ensembles to reduce the sampling uncertainty, but only one representation of possible sea-surface temperature change. UKCP18 is dominated by a single model, but one which is from the latest generation of models and is higher resolution than most models in the other MIPs,

potentially capturing phenomena not resolved by coarser GCMs. Finally HELIX contains two high-resolution atmospheric models, and spans a range of possible sea-surface temperature trends estimated from different CMIP5 models. These factors contribute to different effective degrees of freedom and reliability of each ensemble (e.g., Yokohata et al. 2013), resulting in different estimates of uncertainty and ranges of possible future changes.

To help identify the most likely future changes, increasing the number of models gives a better idea of all of the possible climate responses. In this method, including more samples in the central estimate reduces the uncertainty by narrowing in on the forced change (where models agree). However this does not necessarily remove the possibility of the true changes being outside our confidence intervals, where there are outlier models. Unless there are physical reasons to exclude a particular outlying model they should still be considered plausible scenarios. We note that the multi-model 'central estimate' changes, represent the mean change in the metrics considered and do not span the full model spread including outliers. For purposes of risk assessments, worst case projections based on the full probability distribution of projections (e.g., Sutton 2019; Quinn et al. 2013), should be used in addition to the 'central estimate'. This can take into account changes in variability and likelihood of particular extreme events occurring, which is important for decision making. We note that combining projections of extremes from atmosphere-only and coupled ocean-atmosphere model activities could be more problematic, as the SST-forced simulations exhibit a smaller range of variability due to sampling a smaller range of possible climate states (Fischer et al. 2018). So a multi-MIP analysis of extreme weather events may benefit from including a method of correcting variability (e.g., Bellprat et al. 2019) or by restricting to similar model configurations (e.g. coupled model only).

In this study, we chose a methodology to produce the multi-model 'central estimates', which does not account for model skill. Models have different biases and skill in representing historical

climate change. SST forced atmospheric models for example generally have lower biases than coupled models (He and Soden 2016), and model developers are constantly working to improve their model's performance which may result in differences between generations of models. Because of this, it may be desirable to weight models, for example on their representation of different aspects of current climate (Sanderson et al. 2017a; Shiogama et al. 2011; Knutti 2010). Including model skill in the analysis could give greater (or lower) weighting to outlying results from models that are better (or worse) at representing a specific phenomena. The approach of considering all models equal is a limitation of our methodology, and exploring this further will add to the conclusions of this study.

In our analyses we consider the projections of each MIP equally plausible when combining their estimates. In reality, the projections of specific MIPs are not equal and will have strengths and weaknesses. However as it is common practice in the scientific literature to base their conclusions on a single MIP, we combine these separate estimates without giving one higher consideration than the others. Separate to how realistic the projections are, there are various inter-dependencies between the MIPs. These can be due to including models with commonalities (e.g. different generations of the same model or different models with shared components) (Knutti et al. 2013). In addition, the HAPPI and HELIX projects use SST projections based on output from CMIP5 and UKCP18 also includes some results from CMIP5. When combining results from different MIPs, adding additional independent data sources should increase the confidence of our projections. However, the presence of common information could narrow the uncertainty range in an unrealistic way by treating data with similar origins as independent sources. As such, the use of the combined central estimate should be used to complement an evaluation of different MIPs rather than replacing such an analysis. A future refinement of the methodology used here could take into account factors such as the inter-dependence of the MIPs, skill of models within the MIPs and abilities of the

MIPs to sample a wide range of plausible future states. We expect that such weighting of MIPs would modify the overall confidence ranges produced by this analysis, however the details of this weighting is beyond the scope of this work.

Another limitation of combining results from different modeling activities is that the results may be harder to interpret. The combined results do not have the same specificity regarding the experimental design as results that are, for example, reflecting the trajectory of a single future scenario. The combined central estimates presented here reflect possible changes to 1.5°C and 2°C, but if there are differences important for policy reasons such as between transient and stabilized climates (e.g., Zappa et al. 2020; King et al. 2020), this may necessitate considering a smaller number of simulations that are relevant to the specific question at hand.

Use of single model large ensembles also has the potential to disentangle the uncertainty due to differences in model responses and experimental design. In Fig. 6, we use two large ensembles to show differences in precipitation response between transient and stabilized climate scenarios. As more of these ensembles become available (e.g., Deser et al. 2020) they will be a valuable tool for comparing results across MIPs with consistent model structures.

This study emphasizes that analyzing precipitation changes using a single MIP does not fully take advantage of previous modeling work. The IPCC AR6 is likely to focus on results from CMIP6 at the expense of previous activities, however this may over-estimate the confidence in precipitation changes. Furthermore, in some cases, using CMIP6 on its own gives different changes compared to other methods used here. Combining information from different modeling activities will improve our understanding of confidence in the changes and where the uncertainty lies, and should be adopted when formulating climate policy.

Data availability statement. Climate simulations used in this study are freely accessible to the public with the exceptions of the HELIX data which is available by request.

Acknowledgments. We acknowledge the World Climate Research Programme, which, through its Working Group on Coupled modeling, coordinated and promoted CMIP5 and CMIP6. We thank the climate modeling groups for producing and making available their model output, the Earth System Grid Federation (ESGF) for archiving the data and providing access, and the multiple funding agencies who support CMIP5, CMIP6 and ESGF.

This research used science gateway resources of the National Energy Research Scientific Computing Center, a DOE Office of Science User Facility supported by the Office of Science of the U.S. Department of Energy under Contract No. DE-AC02-05CH11231. The work of RAB and the production of the HELIX simulations were supported by the European Union Seventh Framework Programme FP7/2007–2013 under grant agreement no 603864 (HELIX: "High-End cLimate Impacts and eXtremes", www.helixclimate.eu) and the UK BEIS/Defra Met Office Hadley Centre Climate Programme (GA01101). PDB is supported by a Royal Society Wolfson Research Merit Award. CH acknowledges the Natural Environment Research Council National Capability award to the Centre for Ecology and Hydrology. ADK is funded by the Australian Research Council (DE180100638).

APPENDIX

Climate model data-sets

a. CMIP5

The Coupled Modeling Intercomparison Project, Phase 5 (CMIP5) (Taylor et al. 2012) is the modeling effort used as the basis for the IPCC AR5. It involved a large number (> 30) of different

climate models and in this study we use the historical simulations and future scenarios following the Representative Concentration Pathways (RCPs) specified in the CMIP5 protocol. The models included in this study and the number of ensemble members used for each level of global warming are given in Table A1.

b. CMIP6

The Coupled Modeling Intercomparison Project, Phase 6 (CMIP6) (Eyring et al. 2016) is designed to inform the IPCC Sixth Assessment Report. At the time of writing, new simulations from CMIP6 are still being added to the CMIP6 archive. So estimates of change using this data-set may change as additional models are included. In this study we use the historical simulations and future scenarios following Shared Socioeconomic Pathways (SSPs), from the ScenarioMIP activity (O'Neill et al. 2016). The models included in this study and the number of ensemble members used for each level of global warming are given in Table A2.

c. HAPPI

Simulations run for the Half a degree Additional warming, Prognosis and Projected Impacts project (HAPPI) (Mitchell et al. 2017) are 10 year atmosphere-only climate simulations, forced by sea-surface temperatures (SSTs), sea-ice concentration (SIC) and green-house gas concentrations. The presend day period used in HAPPI is 2006–2015, and uses observed SSTs from the OSTIA observational data-set (Donlon et al. 2012). SSTs from CMIP5 model output are used to estimate the future scenarios corresponding to 1.5°C and 2°C global warming. These simulations are targeted to simulate 1.5°C and 2°C warming, so do not require calculation of time-slices.

Large ensembles were produced by running simulations with different initial condition perturbations. The models included in this study and the number of ensemble members used for each level of global warming are given in Table A3.

d. UKCP18

The 2018 UK Climate Projections (UKCP18) global 60km product (Murphy et al. 2019) was used. This consists of a perturbed physics ensemble of 15 HadGEM3-GC3.05 simulations supplemented by 13 CMIP5 projections, each from different models. These simulations follow the RCP8.5 protocol and time-slices for specific levels of warming have been extracted using the same method as per CMIP5 and CMIP6. This data-set was developed to make use of the higher resolution and more complex physics of HadGEM3-GC3.05 than is available in current MIPs.

e. HELIX

HELIX (High-End cLimate Impacts and eXtremes) is a European Commission funded initiative to produce climate projections using high resolution global atmospheric models. Two models were used: EC-EARTH3-HR, with resolution nominally corresponding to 40km and HadGEM3-A Global Atmosphere (GA) 3.0 model (Betts et al. 2018) at a resolution of 60km. These models were forced by SSTs from 6 or 7 different CMIP5 models for HadGEM3 and EC-EARTH-HR respectively. This allows the atmospheric models to sample a range of different ocean responses.

The simulations were run from the historical period to 2100 using the RCP8.5 scenario. See Wyser et al. (2017) for details of these simulations. Time-slices were chosen for the 1.5°C and 2°C specific warming levels as specified in the HELIX methodology. We chose to use the current climate time-slice as 2000-2020. This is because specific warming levels less than 1.5°C were not defined in the HELIX methodology.

f. CESM large ensemble and low warming simulations

The Community Earth System Model (CESM) has computed a large ensemble ('CESM-LE') of historical and RCP8.5 simulations following the CMIP5 protocol (Kay et al. 2015). In addition, targeted low warming simulations with the same model ('LowWarm') (Sanderson et al. 2017b), were run from 2006-2100. These simulations use tailored emissions pathways to achieve stabilized climate at 1.5°C or 2°C by 2100. The LowWarm simulations branch from a subset (11) of the CESM-LE historical simulations, so can be considered as continuous simulations from 1920–2100.

For calculating the warming since pre-industrial, we note that one of the historical simulations starts at 1850, but the rest start at 1920, so a base period of 1920–1940 was used to calculate the warming since pre-industrial for each simulation. To keep consistency with other data-sets, the warming between 1861–1900 and 1920–1940 from the longer simulation was added to the warming amount relative to the 1920–1940 base period.

Comparing the CESM-LE and LowWarm simulations allows a quantification of the difference caused by the experimental design for a given model structure.

g. CanESM2 large ensembles

The CanESM2 model (Arora et al. 2011) also has a large ensemble of coupled model simulations. These were created by branching from the CMIP5 historical simulations at 1950, with different simulations produced by using different random number seed values in the cloud parameterization (Kirchmeier-Young et al. 2017). Historical simulations were run from 1950–2005, and then continued using RCP8.5 forcing from 2006–2100. To determine the global mean warming since pre-industrial conditions for the CanESM2 large ensemble simulations, these simulations were extended back to 1861 by the corresponding CMIP5 simulations.

The atmospheric component of the CanESM2 model was also used in the HAPPI project. This allows an estimate of influence of the experimental design between HAPPI and CMIP5, although this also includes the difference between a coupled atmosphere-ocean model and an atmosphere-only model.

References

Allen, M. R., and W. J. Ingram, 2002: Constraints on future changes in climate and the hydrologic cycle. *Nature*, **419 (6903)**, 228–232, doi:10.1038/nature01092.

Arora, V. K., and Coauthors, 2011: Carbon emission limits required to satisfy future representative concentration pathways of greenhouse gases. *Geophys. Res. Lett.*, **38 (5)**, doi: 10.1029/2010GL046270.

Bellprat, O., V. Guemas, F. Doblas-Reyes, and M. G. Donat, 2019: Towards reliable extreme weather and climate event attribution. *Nature communications*, **10 (1)**, 1–7, doi: 10.1038/s41467-019-09729-2.

Betts, R. A., and Coauthors, 2018: Changes in climate extremes, fresh water availability and vulnerability to food insecurity projected at 1.5°C and 2°C global warming with a higher-resolution global climate model. *Philos Transact A Math Phys Eng Sci*, **376 (2119)**, doi:10.1098/rsta.2016.0452.

Brunner, L., A. G. Pendergrass, F. Lehner, A. L. Merrifield, R. Lorenz, and R. Knutti, 2020: Reduced global warming from CMIP6 projections when weighting models by performance and independence. *Earth System Dynamics Discussions*, **2020**, 1–23, doi:10.5194/esd-2020-23, in review.

626 Cochran, W. G., 1937: Problems arising in the analysis of a series of similar experiments. *Suppl.*
627 *J. Royal Stat. Soc.*, **4 (1)**, 102–118, URL http://www.jstor.org/stable/2984123.

628 Collins, M., and Coauthors, 2013: *Long-term Climate Change: Projections, Commitments and*
629 *Irreversibility. In: Climate Change 2013: The Physical Science Basis. Contribution of Working*
630 *Group I to the Fifth Assessment Report of the Intergovernmental Panel on Climate Change*.
631 Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA.

632 DerSimonian, R., and N. Laird, 1986: Meta-analysis in clinical trials. *Control. Clin. Trials*, **7 (3)**,
633 177–188, doi:10.1016/0197-2456(86)90046-2.

634 Deser, C., A. Phillips, V. Bourdette, and H. Teng, 2012: Uncertainty in climate change projections:
635 the role of internal variability. *Clim. Dyn.*, **38 (3)**, 527–546, doi:10.1007/s00382-010-0977-x.

636 Deser, C., and Coauthors, 2020: Insights from earth system model initial-condition large ensembles
637 and future prospects. *Nature Climate Change*, 1–10, doi:10.1038/s41558-020-0731-2.

638 Donlon, C., M. Martin, J. Stark, J. Roberts-Jones, E. Fiedler, and W. Wimmer, 2012: The
639 Operational Sea Surface Temperature and Sea Ice Analysis (OSTIA) system. *Remote Sens.*
640 *Environ.*, **116 (Supplement C)**, 140–158, doi:https://doi.org/10.1016/j.rse.2010.10.017, URL
641 http://www.sciencedirect.com/science/article/pii/S0034425711002197.

642 Eyring, V., S. Bony, G. A. Meehl, C. A. Senior, B. Stevens, R. J. Stouffer, and K. E. Taylor, 2016:
643 Overview of the Coupled Model Intercomparison Project Phase 6 (CMIP6) experimental design
644 and organization. *Geosci. Model Dev.*, **9 (5)**, 1937–1958, doi:10.5194/gmd-9-1937-2016, URL
645 https://www.geosci-model-dev.net/9/1937/2016/.

Fischer, E. M., U. Beyerle, C. F. Schleussner, A. D. King, and R. Knutti, 2018: Biased Estimates of Changes in Climate Extremes From Prescribed SST Simulations. *Geophysical Research Letters*, **45 (16)**, 8500–8509, doi:10.1029/2018GL079176.

Fischer, E. M., J. Sedláček, E. Hawkins, and R. Knutti, 2014: Models agree on forced response pattern of precipitation and temperature extremes. *Geophysical Research Letters*, **41 (23)**, 8554–8562, doi:10.1002/2014GL062018.

Good, P., B. B. Booth, R. Chadwick, E. Hawkins, A. Jonko, and J. A. Lowe, 2016: Large differences in regional precipitation change between a first and second 2 k of global warming. *Nat. Commun.*, **7**, 13 667.

Hawkins, E., and R. Sutton, 2009: The potential to narrow uncertainty in regional climate predictions. *Bull. Am. Meteorol. Soc.*, **90 (8)**, 1095–1108, doi:10.1175/2009BAMS2607.1.

Hawkins, E., and R. Sutton, 2011: The potential to narrow uncertainty in projections of regional precipitation change. *Clim. Dyn.*, **37 (1)**, doi:10.1007/s00382-010-0810-6.

He, J., and B. Soden, 2016: The impact of SST biases on projections of anthropogenic climate change: A greater role for atmosphere-only models? *Geophysical Research Letters*, **43 (14)**, 7745–7750, doi:10.1002/2016GL069803.

Held, I. M., M. Winton, K. Takahashi, T. Delworth, F. Zeng, and G. K. Vallis, 2010: Probing the fast and slow components of global warming by returning abruptly to preindustrial forcing. *Journal of Climate*, **23 (9)**, 2418–2427, doi:10.1175/2009JCLI3466.1.

IPCC, 2018: *Impacts of 1.5°C Global Warming on Natural and Human Systems. In: Global Warming of 1.5°C. An IPCC Special Report on the impacts of global warming of 1.5°C above pre-industrial levels and related global greenhouse gas emission pathways, in the context of*

strengthening the global response to the threat of climate change, sustainable development, and efforts to eradicate poverty. In Press.

Iturbide, M., and Coauthors, 2020: An update of IPCC climate reference regions for subcontinental analysis of climate model data: Definition and aggregated datasets. *Earth System Science Data Discussions*, **2020**, 1–16, doi:10.5194/essd-2019-258.

James, R., R. Washington, C.-F. Schleussner, J. Rogelj, and D. Conway, 2017: Characterizing half-a-degree difference: a review of methods for identifying regional climate responses to global warming targets. *Wiley Interdiscip. Rev. Clim. Change*, **8 (2)**, e457–n/a, doi:10.1002/wcc.457.

Kay, J. E., and Coauthors, 2015: The Community Earth System Model (CESM) Large Ensemble Project: A Community Resource for Studying Climate Change in the Presence of Internal Climate Variability. *Bull. Am. Meteorol. Soc.*, **96 (8)**, 1333–1349, doi:10.1175/BAMS-D-13-00255.1.

King, A. D., D. J. Karoly, and B. J. Henley, 2017: Australian climate extremes at 1.5 °C and 2 °C of global warming. *Nat. Clim. Change*, **7 (6)**, 412–416.

King, A. D., R. Knutti, P. Uhe, D. M. Mitchell, S. C. Lewis, J. M. Arblaster, and N. Freychet, 2018: On the Linearity of Local and Regional Temperature Changes from 1.5°C to 2°C of Global Warming. *Journal of Climate*, **31 (18)**, 7495–7514, doi:10.1175/JCLI-D-17-0649.1.

King, A. D., T. P. Lane, B. J. Henley, and J. R. Brown, 2020: Global and regional impacts differ between transient and equilibrium warmer worlds. *Nature Climate Change*, **10 (1)**, 42–47, doi:10.1038/s41558-019-0658-7.

Kirchmeier-Young, M. C., F. W. Zwiers, and N. P. Gillett, 2017: Attribution of Extreme Events in Arctic Sea Ice Extent. *J. Clim.*, **30 (2)**, 553–571, doi:10.1175/JCLI-D-16-0412.1.

689 Knutti, R., 2010: The end of model democracy? *Climatic Change*, **102 (3)**, 395,404, doi:
690 10.1007/s10584-010-9800-2.

691 Knutti, R., R. Furrer, C. Tebaldi, J. Cermak, and G. A. Meehl, 2010: Challenges in Com-
692 bining Projections from Multiple Climate Models. *J. Clim.*, **23 (10)**, 2739–2758, doi:
693 10.1175/2009JCLI3361.1.

694 Knutti, R., D. Masson, and A. Gettelman, 2013: Climate model genealogy: Generation cmip5 and
695 how we got there. *Geophysical Research Letters*, **40 (6)**, 1194–1199, doi:10.1002/grl.50256.

696 Knutti, R., and J. Sedláček, 2013: Robustness and uncertainties in the new CMIP5 climate model
697 projections. *Nature Climate Change*, **3 (4)**, 369–373, doi:10.1038/nclimate1716.

698 Lehner, F., C. Deser, N. Maher, J. Marotzke, E. M. Fischer, L. Brunner, R. Knutti, and E. Hawkins,
699 2020: Partitioning climate projection uncertainty with multiple large ensembles and CMIP5/6.
700 *Earth System Dynamics*, **11 (2)**, 491–508, doi:10.5194/esd-11-491-2020.

701 Manabe, S., R. J. Stouffer, M. J. Spelman, and K. Bryan, 1991: Transient responses of a coupled
702 ocean–atmosphere model to gradual changes of atmospheric co2. part i. annual mean response.
703 *Journal of Climate*, **4 (8)**, 785–818, doi:10.1175/1520-0442(1991)004<0785:TROACO>2.0.
704 CO;2.

705 Merrifield, A. L., L. Brunner, R. Lorenz, I. Medhaug, and R. Knutti, 2020: An investigation
706 of weighting schemes suitable for incorporating large ensembles into multi-model ensembles.
707 *Earth System Dynamics*, **11 (3)**, 807–834, doi:10.5194/esd-11-807-2020.

708 Mitchell, D., R. James, P. M. Forster, R. A. Betts, H. Shiogama, and M. Allen, 2016: Realizing the
709 impacts of a 1.5 °C warmer world. *Nat. Clim. Change*, **6 (8)**, 735–737.

Mitchell, D., and Coauthors, 2017: Half a degree additional warming, prognosis and projected impacts (HAPPI): background and experimental design. *Geosci. Model Dev.*, **10 (2)**, 571–583, doi:10.5194/gmd-10-571-2017.

Murphy, J., and Coauthors, 2019: UKCP18 Land Projections: Science Report. Tech. rep., UK Met Office. URL https://www.metoffice.gov.uk/pub/data/weather/uk/ukcp18/science-reports/UKCP18-Land-report.pdf.

O'Neill, B. C., and Coauthors, 2016: The Scenario Model Intercomparison Project (ScenarioMIP) for CMIP6. *Geosci. Model Dev.*, **9 (9)**, 3461–3482, doi:10.5194/gmd-9-3461-2016.

Pendergrass, A. G., F. Lehner, B. M. Sanderson, and Y. Xu, 2015: Does extreme precipitation intensity depend on the emissions scenario? *Geophys. Res. Lett.*, **42 (20)**, 8767–8774, doi:10.1002/2015GL065854.

Quinn, N., P. D. Bates, and M. Siddall, 2013: The contribution to future flood risk in the Severn Estuary from extreme sea level rise due to ice sheet mass loss. *J. Geophys. Res. Oceans*, **118 (11)**, 5887–5898, doi:10.1002/jgrc.20412.

Sanderson, B. M., M. Wehner, and R. Knutti, 2017a: Skill and independence weighting for multi-model assessments. *Geosci. Model Dev.*, **10 (6)**, 2379–2395, doi:10.5194/gmd-10-2379-2017.

Sanderson, B. M., and Coauthors, 2017b: Community climate simulations to assess avoided impacts in 1.5 and 2 °C futures. *Earth Syst. Dyn.*, **8 (3)**, 827–847, doi:10.5194/esd-8-827-2017.

Seneviratne, S., and Coauthors, 2018: The many possible climates from the Paris Agreement's aim of 1.5 °C warming. *Nature*, **558 (7708)**, 41–49, doi:10.1038/s41586-018-0181-4.

Sherwood, S. C., S. Bony, and J.-L. Dufresne, 2014: Spread in model climate sensitivity traced to atmospheric convective mixing. *Nature*, **505 (7481)**, 37–42, doi:10.1038/nature12829.

Shiogama, H., S. Emori, N. Hanasaki, M. Abe, Y. Masutomi, K. Takahashi, and T. Nozawa, 2011: Observational constraints indicate risk of drying in the amazon basin. *Nature Communications*, **2**, 253, doi:10.1038/ncomms1252.

Sutton, R. T., 2019: Climate Science Needs to Take Risk Assessment Much More Seriously. *Bulletin of the American Meteorological Society*, **100 (9)**, 1637–1642, doi:10.1175/BAMS-D-18-0280.1.

Taylor, K., R. L. Stouffer, and G. A. Meehl, 2012: An overview of CMIP5 and the experiment design. *Bull. Amer. Meteor. Soc.*, **93**, 485–498, doi:10.1175/BAMS-D-11-00094.1.

Tebaldi, C., J. M. Arblaster, and R. Knutti, 2011: Mapping model agreement on future climate projections. *Geophys. Res. Lett.*, **38 (23)**, doi:10.1029/2011GL049863.

Tokarska, K. B., M. B. Stolpe, S. Sippel, E. M. Fischer, C. J. Smith, F. Lehner, and R. Knutti, 2020: Past warming trend constrains future warming in cmip6 models. *Science Advances*, **6 (12)**, doi:10.1126/sciadv.aaz9549.

Uhe, P., D. Mitchell, P. Bates, C. Sampson, A. Smith, and A. Islam, 2019: Enhanced flood risk with 1.5°C global warming in the Ganges-Brahmaputra-Meghna basin. *Environ. Res. Lett.*, **14 (7)**, 074 031, doi:10.1088/1748-9326/ab10ee.

Wallemacq, P., and R. House, 2018: Economic losses, poverty & disasters: 1998-2017. Tech. rep., Centre for Research on the Epidemiology of Disasters (CRED) and United Nations Office for Distaster Risk Reduction (UNISDR), 31 pp. URL https://www.preventionweb.net/go/61119.

Wyser, K., G. Strandberg, J. Caesar, and L. Gohar, 2017: Documentation of changes in climate variability and extremes simulated by the HELIX AGCMs at the 3 SWLs and comparison to changes in equivalent SST/SIC low-resolution CMIP5 projections. Tech. rep. URL https://helixclimate.eu/working-packages/high-resolution-timeslices-and-regional-downscaling-wp3.

Yokohata, T., and Coauthors, 2013: Reliability and importance of structural diversity of climate model ensembles. *Climate Dynamics*, **41 (9)**, 2745–2763, doi:10.1007/s00382-013-1733-9.

Zappa, G., P. Ceppi, and T. G. Shepherd, 2020: Time-evolving sea-surface warming patterns modulate the climate change response of subtropical precipitation over land. *Proceedings of the National Academy of Sciences*, **117 (9)**, 4539–4545, doi:10.1073/pnas.1911015117.

Zelinka, M. D., T. A. Myers, D. T. McCoy, S. Po-Chedley, P. M. Caldwell, P. Ceppi, S. A. Klein, and K. E. Taylor, 2020: Causes of higher climate sensitivity in cmip6 models. *Geophysical Research Letters*, **47 (1)**, e2019GL085 782, doi:10.1029/2019GL085782.

LIST OF FIGURES

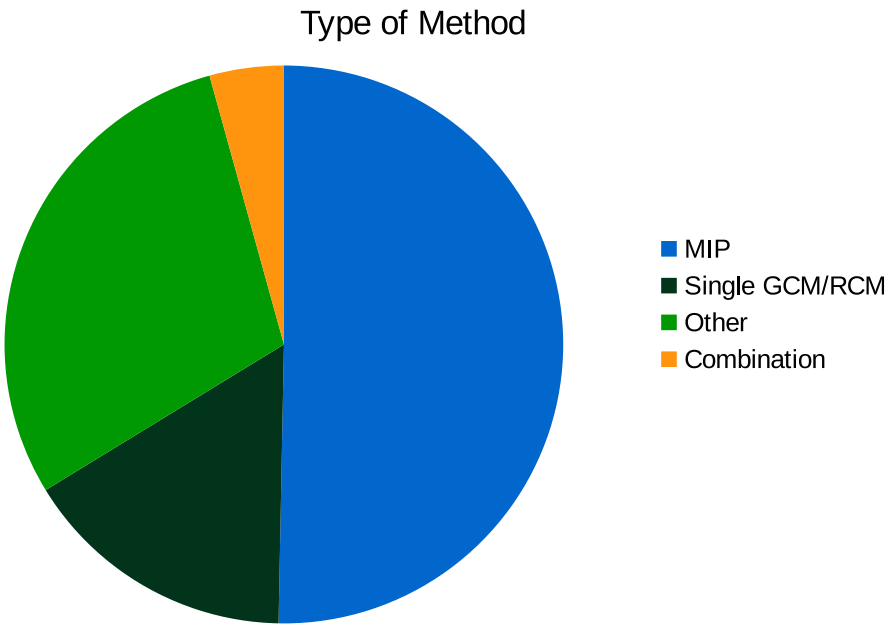


FIG. 1. **Categorization of projections in IPCC SR1.5 impacts studies**

Categorization of methods used by papers in the IPCC SR1.5 impacts chapter (chapter 3), considering 163 studies. 'MIP' includes simulations from the CMIP5, CMIP3, CORDEX, HAPPI and HELIX modeling protocols. 'Other' refers to methods which do not directly use GCMs. Note that some studies, e.g. using climate emulators, may be based around GCM results indirectly.

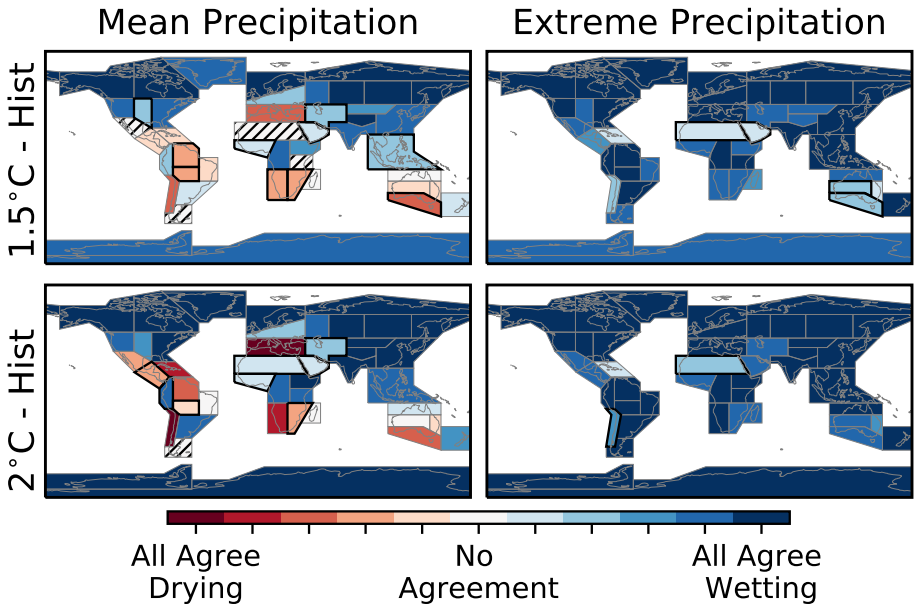


FIG. 2. **Agreement in projections: mean and extreme precipitation**

Agreement between modeling activities (CMIP5, CMIP6, HAPPI, UKCP18, HELIX) in a significant change for mean and extreme precipitation. Changes are calculated between time-slices at specific warming levels: 1.5°C vs Hist (upper), 2°C vs Hist (lower). 'Hist' refers to current climate (see section 2c). If two methods show opposing changes, this is assigned no agreement and the region is hatched. Changes are calculated for regional means over AR5 reference regions, for yearly mean precipitation (left) and yearly maximum of 5 day precipitation (Extreme precipitation, right). Regions with bold outlines are where CMIP6 agrees with at most one other method about the sign and significance of the change.

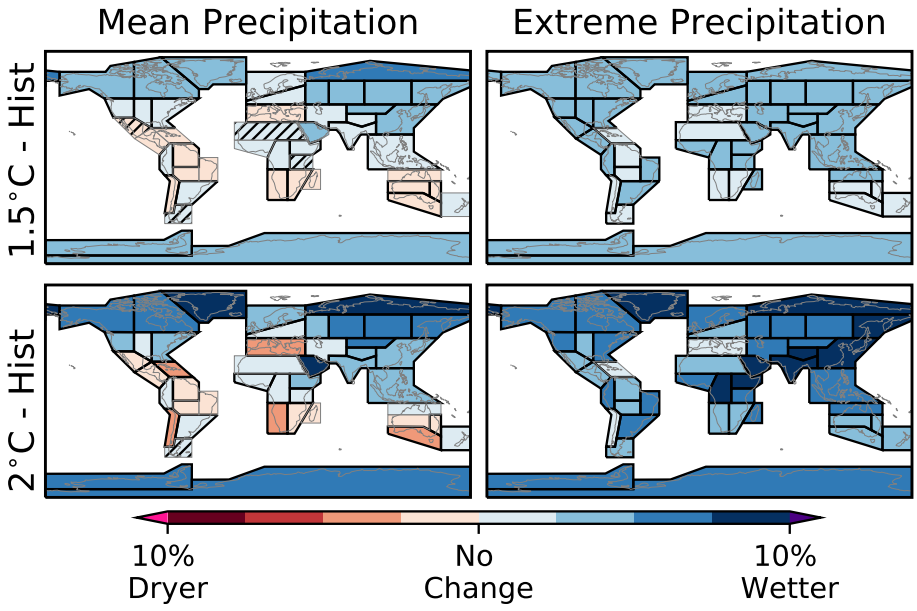


FIG. 3. **Multi-method projections: mean and extreme precipitation**

Combined central estimate of changes across 5 modeling activities (CMIP5, CMIP6, HAPPI, UKCP18, HELIX). Changes are calculated between time-slices at specific warming levels: 1.5°C vs Hist (upper), 2°C vs Hist (lower). 'Hist' refers to current climate (see section 2c). Bold region outlines indicate significance in the change, e.g. where the combined central estimate 5-95% confidence interval does not include zero. Hatching is used to indicate where two methods show opposing significant changes. Changes are calculated for regional means over AR5 reference regions, for yearly mean precipitation (left) and yearly maximum of 5 day precipitation (Extreme precipitation, right).

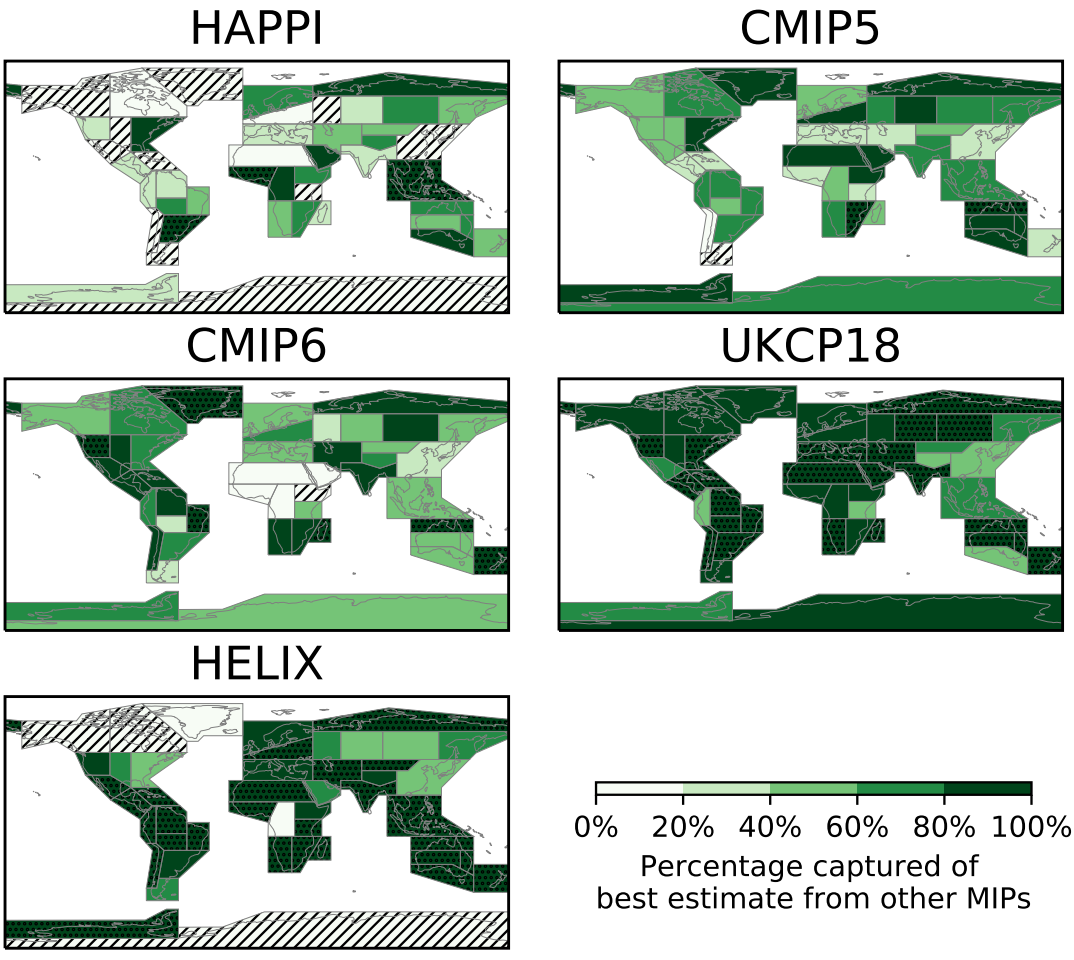


FIG. 4. **Amount of uncertainty from the combined estimates captured by the uncertainty of specific MIPs**

For mean precipitation, 1.5°C - current climate, and each MIP, this shows the percentage coverage by its confidence interval, of the combined central estimate interval from the other four MIPs. Here 0% (hashed regions) indicates that the MIP is in complete disagreement with the combined estimate from the other MIPs. 100% (stippled regions) indicates that the confidence interval of the MIP completely encompasses the combined confidence range from the other MIPs. The confidence intervals are the 5-95% range for the change in mean precipitation between current climate and 1.5°C warming.

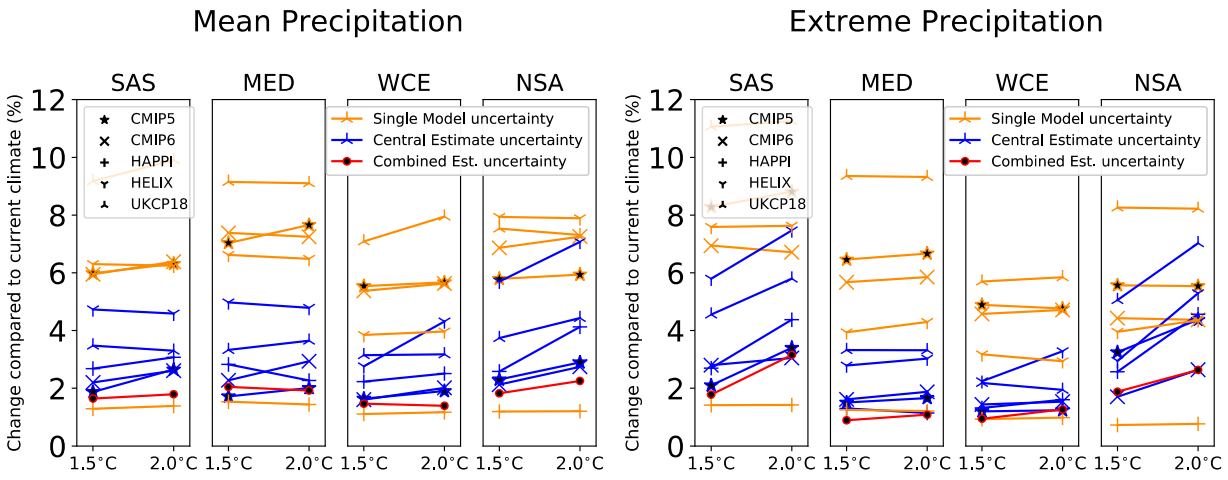


FIG. 5. **Partitioning of Uncertainty: mean and extreme precipitation**

Plots showing estimates in uncertainty in changes of precipitation between current climate and 1.5°C or 2°C climates, for four regions. Orange markers give the median 'single model uncertainty', given as the (5–95% confidence interval in the changes from bootstrap re-sampling) for a particular MIP. Blue markers give estimates of the 5–95% confidence interval of the 'central estimate' change for each MIP. The red dots show the 5–95% confidence interval range of the 'Combined central estimate' changes. Regions are SAS: South Asia, MED: Mediterranean, WCE: western Central Europe, NSA: northern South America

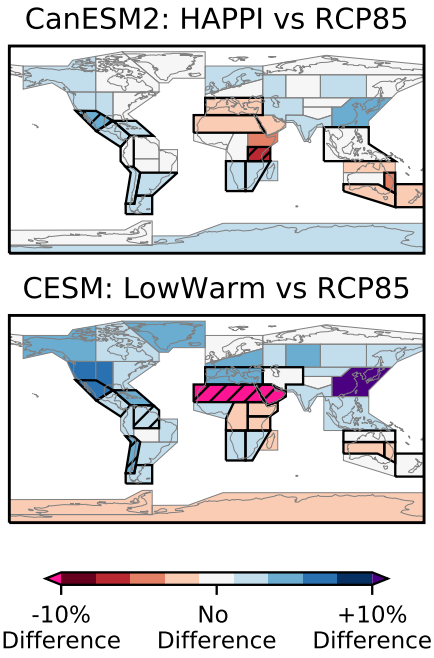
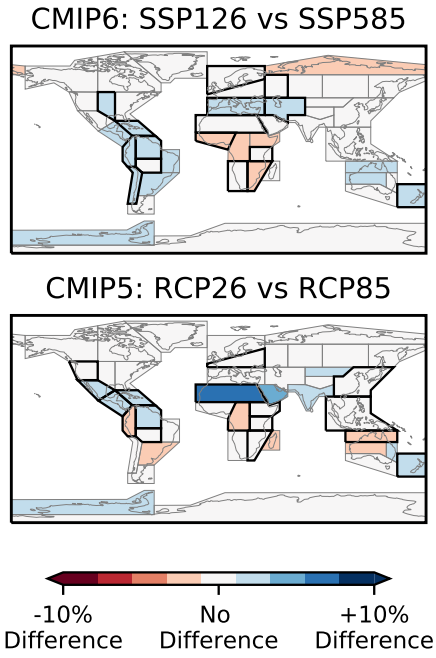


FIG. 6. **Experimental design difference for mean precipitation changes between current climate and 1.5°C**

Differences between changes from the same model with different experimental designs. The differences are calculated for percentage changes in mean precipitation, comparing 1.5°C to current climate. The top panel shows results using the CanESM2 model: HAPPI simulations (using atmospheric component CanAM4) vs RCP8.5 simulations (CanESM2 large ensemble). The lower panel shows results using the CESM-CAM5 model: CESM-CAM5: Low Warming simulations (LowWarm) compared to RCP8.5 simulations. Regions with bold outlines are where the significance of the change is different between the experimental designs. Hatched regions are where the different experimental designs result in opposite significant changes.



867 FIG. 7. **Emission scenario difference for mean precipitation changes between current climate and 1.5°C**
868 Differences between changes from the same modeling activity, comparing high and low emissions scenarios.
869 These show SSP126 vs SSP585 for CMIP6 (top) and RCP26 vs RCP85 for CMIP5 (bottom). The differences are
870 calculated for percentage changes in mean precipitation, comparing 1.5°C to current climate. Regions with bold
871 outlines are where the significance or sign of the change is different between the scenarios.